UNCLASSIFIED

AD 410446

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

N-63-4-3

Western Management Science Institute
University of California   •   Los Angeles

University of California

Los Angeles

Western Management Science Institute and

Department of Preventive Medicine and Public Health

Working Paper No. 32

June 4, 1963

OPTIMAL POLICIES FOR A CLASS OF SEARCH AND EVALUATION PROBLEMS

by

J. MacQueen

OPTIMAL POLICIES FOR A CLASS OF SEARCH AND EVALUATION PROBLEMS[1]

J. MacQueen[2]

University of California, Los Angeles

1. Introduction

Not infrequently, the decision-making process goes something like this: there is first given some sort of initiating problem or goal. This leads to exploration, or search, for possible solutions. As various possibilities are discovered, they are evaluated tentatively and some idea of their worth as solutions to the initiating problem is obtained. On this basis, it may be decided to accept a certain possibility without further consideration, or it may be decided to evaluate a certain possibility more carefully, that is, to experiment or test, or, it may be decided to seek other possibilities, that is, to continue exploring. However, after varying amounts of exploration alternating with various amounts of evaluation, an acceptable possibility is eventually located and the process is stopped, at least as regards the particular problem at hand.

In this paper, we present a mathematical model for decision making as viewed in this light, i.e., as a more or less sequential process which involves constant alternation between exploratory and evaluative operations until a satisfactory possibility is located. We

then describe the theoretical optimal behavior in the context of the model. The optimal behavior is of possible normative interest, and is of use in experimental studies of decision making where it is desirable to compare actual behavior with various kinds of theoretical decision rules.

The features of decision making and behavior in general noted above have been described by a number of writers. Among others, E. C. Tolman, John Dewey, and H. A. Simon have paid special attention to the activity of arriving at a course of action through active exploration and evaluation, and more or less formal models relevant to this kind of phenomena have been described by several persons, e.g., Stigler (8), Ashby(1), and Toda(9), as well as Simon(7). The model described here supplements this work by making possible explicit description of the optimal balancing of exploratory and evaluative activity within a well-defined context. We give up a certain amount of the generality inherent in some of the above models in order to gain a measure of precision.

It is possible to give many instances of group and individual behavioral sequences, of the general sort alluded to above, in which decisions vis-a-vis exploratory and evaluational activities play a central role. These decisions appear to permeate human experience and behavior, and it seems likely that differences in policy with respect to such decisions, often operating in an intuitive and unconsidered fashion, lie at the basis of important differences in individual and organizational performance. For this reason alone, it seems worthwhile to attempt to obtain an explicit and thorough understanding of the problem under various sets of limited conditions such as are considered here.

Section 2 below describes the formal model and Section 3 presents the mathematical results characterizing the optimal policy. The main result is that the solutions to certain given systems of equations uniquely characterize the optimal policy. The proofs of these results, which are not difficult, are given in Section 4.

In Section 5, the model is applied to the case where it is desired to carry on many "search and evaluation" processes of the same sort, subject to a constraint on the total expenditure for conducting all of them. This version of the model may be of value in practical selection problems where many objects must be picked out. Section 6 points out some aspects of the problem of obtaining numerical solutions when the distributions governing the search and evaluation process have the form of the joint normal.

2. The Model

Let $(X_1, Y_1, V_1), (X_2, Y_2, V_2), \ldots$ be a sequence of independent, real-valued triples, with known, common joint distribution H. The decision maker first pays an amount $c_S > 0$, called the search cost, which gives him an opportunity to choose a possibility whose worth is given by V_1. However, the decision maker is told only X_1, which, because of the joint distribution H, gives him some information about V_1. At this point he may either stop, taking V_1 as his reward, continue, in which case he has a chance at V_2 and learns X_2, or take an action referred to as a test, which costs $c_T > 0$ and enables him to learn Y_1, thus gaining more information about V_1. In the latter case, having observed Y_1, again he may stop, receiving V_1, or continue, getting a chance at V_2 and learning X_2, still at a cost c_S. However, this entails permanent loss of the option of

taking V_1, as does the decision to continue directly after observing X_1. Once X_2 is observed, these same possibilities are available, and so on. Thus, after continuing n times, X_n is known and the deicion maker can stop, taking V_n, continue on to learn X_{n+1}, or test, observing Y_n, and then again either stop with V_n or continue. Continuing always results in loss of option. The cost* of continuing is always c_S and the cost of a test is always c_T. The problem is, when to continue, when to test, and when to stop, in order to maximize the expected net return.

If the opportunity to test is eliminated, the resulting problem is essentially identical to one mentioned illustratively by MacQueen and Miller(6) and Chow and Robbins(3), except that these writers permitted the decision maker to return to an earlier opportunity if he desired. However, the option of returning to an earlier possibility is never used, and a little consideration shows that our model is not complicated in any important way by permitting complete option. Because of the independence and the fact that H, c_S, and c_T are known and constant, the expectation about the future is constant. If the optimal future is good enough to lead one to pass up an opportunity to test, or having tested, an opportunity to stop, it will always be so. Thus, options on these opportunities will never be used.

The analysis is based heavily on the apriori distribution of the outcome of testing and then stopping given the information X; that is,

---

*We can interpret the search cost c_S as including a certain amount of expense entailed in acquiring the preliminary information X, as well as the cost of producing the possibility.

the distribution of $Z = E(V|X,Y)$ when X is known.[3] This distribution,
which may be obtained from H, is represented by $F(z|x) = P[Z \leq z|X=x]$.
We suppose that Z has a density $f(z|x)$ for each x and that x
itself has a density $f(x)$. The distribution function of X is $F(x)$.
We assume that $f(z|x)$ and $f(x)$ are positive for all z, and all x,
and that $E(V) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} zf(z|x)dzf(x)dx$ exists and is finite. Further-
more, $F(z|x)$ is assumed to be differentiable with respect to x uni-
formly in z.

The above assumptions are more or less regularity assumptions,
which enable certain mathematical tools to be applied. More important,
we will require $F(z|x)$ to satisfy the following two conditions:

C_1, $\partial/\partial x F(z|x) < 0$, and C_2, $E(V|X = x) = \int_{-\infty}^{\infty} zf(z|x)dz \rightarrow \pm \infty$ as
$x \rightarrow \pm \infty$, and there exist functions α, β, with $\alpha(x) \rightarrow \infty$ as $x \rightarrow \infty$
and $\beta(x) \rightarrow -\infty$ as $x \rightarrow -\infty$, such that for arbitrary positive δ_i,
$i = 1,2,3,4$, and all x sufficiently large,

(1a) $\quad F(\alpha(x)|x) \leq \delta_1$ and $\int_{-\infty}^{\alpha(x)} |z| f(z|x)dz \leq \delta_2$,

while for all x sufficiently small,

(1b) $\quad 1 - F(\beta(x)|x) \leq \delta_3$ and $\int_{\beta(x)}^{\infty} |z| f(z|x)dz \leq \delta_4$.

When C_1 obtains, we will say $F(z|x)$ is stochastically
ordered in x. For some pertinent remarks on this concept, see Karlin
(3, p.234) and Lehmann(4,p. 73). The condition is satisfied for many
families or distributions, including the normal and any family in which
x corresponds to a location parameter; that is, where $F(z|x) = G(z-x)$

---

[3]In fact, for purposes of determining the optimal policy, the random variables
Y and V can be dispensed with altogether, and the treatment be based on Z
and X alone, or, as is done in Section 3, on Z and the random variable
$E(V|X)$. However, in the psychological studies of decision making for which
this model was devised, the subject may be required to learn the relation be-
tween the random variables or "cues" Y and X, and the random outcome V,
and it is convenient to have the model formulated in these terms.

for some fixed distribution G.

Condition C_2 is a mathematically convenient way of insuring that the mass in $F(z|x)$ follows the mean in a certain sense. Roughly speaking, the condition implies that for a given possibility, if the information $X = x$ is sufficiently favorable, testing is not likely to disclose that in actuality the possibility is poor, and if the information is sufficiently unfavorable, testing is not likely to disclose that the possibility is actually good. The mathematical function of this condition is seen explicitly in the proof of Lemma 2 below.

We assume there exists an optimal stationary policy, that is, a policy which depends only on the variable information available at each stage, X after continuing, or (X,Y) after testing, and which in fact achieves the least upper bound, assumed finite, with respect to the class of all policies. With this assumption in mind, the term optimal is hereafter used to refer to optimal stationary policies.

When C_1 and C_2 hold, and c_T is positive, the optimal policy takes a particularly simple form. For certain constants, v^*, x^*, y^*, and x^o, either, (i) continue for $X \leq x^*$, and test for $x^* \leq X \leq y^*$, and stop without testing for $X \geq y^*$, and stop or continue after testing depending on whether or not $Z \geq v^*$, or (ii) never test and either continue or stop depending on whether $X \leq x^o$ or $X \geq x^o$. A criterion is given for determining whether or not the optimal policy takes the form (i) or (ii), and equations characterizing the optimal constants v^*, x^*, y^*, and x^o are given.

In the above and throughout the following, a policy is described

in the permissive sense by the largest intervals in which each of the various decisions are allowed under the policy. Whenever a given value of X, or Z, as the case may be, belongs to two such intervals either of the corresponding decisions are permitted. In some cases, a value of X might belong to three intervals, in which case all three of the decisions would be permitted. With this interpretation in mind, we can say the optimal policy is unique, for the various constants v^*, x^*, etc. are uniquely determined.

The method of analysis used here under conditions C_1 and C_2 may be extended to the case where C_2 is dropped, although some care must be used to determine the form of the optimal policy. Because of C_1 there are essentially only five qualitatively distinct kinds of policies: Never test, always test, test for all values of X above some point, test for all values of X below some point, or test only for values of X in an intermediate range. As will be seen below, once the form of the optimal policy is known, explicit equations are easily written for the boundaries of the various intervals. Use of C_2 is one way of narrowing these possibilities down to a fairly interesting case; i.e., where there is either testing in an intermediate range or no testing at all.

Without C_1, the situation becomes much more complicated. It is now possible to have $F(z|x)$ have as much variability as desired for any given range of values of x. Thus, the potential payoff for testing, which roughly speaking depends on this variability, can be very high or low anywhere, and the convenient structure described above cannot be expected to obtain.

3. The Optimal Policy

Let v^* be the expected return from continuing and using the optimal policy thereafter and consider the expressions:

$$(2) \qquad T(x,v^*) = v^*F(v^*|x) + \int_{v^*}^{\infty} zf(z|x)dz - c_T,$$

$$(3) \qquad R(x) = \int_{-\infty}^{\infty} zf(z|x) = E(V|X=x)$$

The function $R(x)$ gives the expected return for stopping without testing when for the possibility at hand $X=x$. The function $T(x,v^*)$ gives the expected return for testing when $X=x$ and proceeding optimally thereafter. This is because after testing, which costs c_T, with probability $F(v^*|x)$, it will turn out that $z \leq v^*$, and it will be optimal to continue, which achieves v^*, hence the term $v^*F(v^*|x)$, and if $z \geq v^*$, it will be optimal to stop, which has expectation z, hence the term $\int_{v^*}^{\infty} zf(z|x)dz$. If v^* were known, the optimal policy after testing would thus be clear, and after continuing the optimal policy could be determined from comparison of v^*, $R(x)$ and $T(x,v^*)$, which for $X=x$ give the returns for the three possible decisions on the assumption that an optimal policy is to be pursued in the future. Selection of the better from among these decisions would thus determine the optimal policy. (We are employing the principle of optimality (2) here.) It remains, then, to determine v^* and show that the optimal policy has the form described in Section 2.

Under C_1 and C_2, $R(x)$ is monotone in x and has the domain $(-\infty,\infty)$. Consequently, we can work with the entirely equivalent random variable R which has value $R(x)$ when $X=x$. This is convenient since

then $E(V|R=r) = r$. Define $x(r)$ by the relation $r = R(x(r))$ and let

$F_1(r) = F(x(r)) = P[R \le r]$ and let $F_1(z|r) = F(z|x(r)) = P[Z \le z|R=r]$.

These distributions have corresponding positive densities $f_1(r)$ and

$f_1(z|r)$. Conditions C_1 and C_2 carry over to the new variable R.

Thus if C_1 holds, $\partial/\partial r F(z|r) < 0$ and if C_2 holds in terms of x,

it will also hold in terms of r with x replaced by $x(r)$ in $\alpha(x)$

and $\beta(x)$, and $f(z|x)$ replaced by $f_1(z|r)$. In these terms, the

expected return for testing when $R=r$ and using the optimal policy

thereafter, becomes

$$(3a) \qquad T_1(r,v^*) = v^* F_1(v^*|r) + \int_{v^*}^{\infty} z f_1(z|r)dz - c_T.$$

Clearly, $T_1(r,v)$ is continuous and differentiable in both r and v.

We will make use of several easy lemmas whose proofs are given

in Section 4. The main point of these is to yield propositions I and II

below.

Lemma 1. Under the stochastic ordering conditions C_1, for every fixed

v, $T_1(r,v)$ is monotone in r with $0 < \partial T_1/\partial r < 1$, and hence for fixed

v there is at most one solution to each of the equations,

$(4) \qquad T_1(r,v) = v,$

$(5) \qquad T_1(s,v) = s,$

and for r^* satisfying (4), $T_1(r,v) < v$ for $r < r^*$, and for s^*

satisfying (5), $T_1(r,v) > r$ for $r < s^*$.

Lemma 2. Under C_2 and with $c_T > 0$, each of the equations (4) and

(5) in fact has at least one finite solution for every v.

Using Lemmas 1 and 2, and by inspection of Figure 1, we obtain
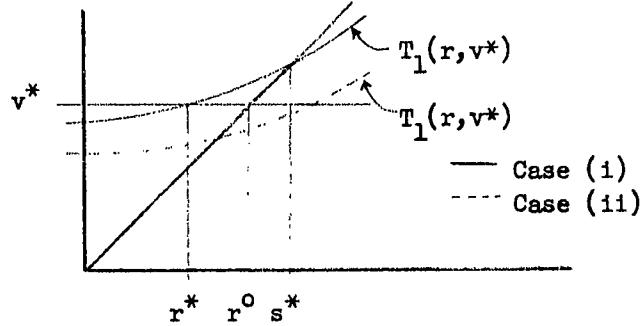
the following:

Figure 1

Lemma 3. Under C_1 and C_2 and with $c_T > 0$, either (i), there is an interval $[r^*,s^*]$ $(r^* \leq s^*)$ such that continuing is optimal given $R = r \leq r^*$, testing is optimal given $R = r\epsilon[r^*,s^*]$, and stopping without testing is optimal given $R=r \geq s^*$, or else (ii), there is no such interval and for some point $r^\circ(=v^*)$, continuing is optimal given $R=r \leq r^\circ$, and stopping without testing is optimal given $R=r \geq r^\circ$.

With Lemma 3 in mind, we easily derive the following:

Proposition I. Suppose that condition (i) of Lemma 3 obtains. Then v^*, together with r^* and s^*, must satisfy the system of equations $\mathfrak{S}$,

(6) $\qquad v = v\mathbb{E}(r) + \int_r^s T_1(t,v)f_1(t)dt + \int_s^\infty tf_1(t)dt - c_s,$

(7) $\qquad T_1(r,v) = v,$

(8) $\qquad T_1(s,v) = s,$

together with the auxiliary condition $r \leq s$. Equation (6) merely equates the expected return from continuing under any policy of the type described under (i), expressed in two different ways. Equation (7) is necessary for the optimal point r^* in as much as from the continuity

of T_1, and lemmas 1 and 2, there is a unique point where $T_1(r,v*) = v*$.
Similarly for equation (8). We note that if (i) fails, this system of
equations cannot be satisfied by the optimal expected return $v*$ and is
meaningless.

A similar argument gives the following:

Proposition II. Under (ii), $v*$ and r^o must satisfy the equations

$$(9) \qquad v = vF_1(r) = \int_r^\infty tf_1(t)dt - c_s$$

and

$$(10) \qquad v = r;$$

that is, $v*(\text{or } r^o)$ must satisfy

$$(9a) \qquad v = vF_1(v) + \int_v^\infty t_1(t)dt - c_s.$$

We are thus in a position to determine $v*$ and the optimal policy
as well, except for two things, the possibility of non-uniqueness of the
solutions to the above systems of equations and the matter of knowing
whether or not (i) or (ii) obtains. Theorems 1, 2, and 3 settle these
questions. The proofs of these theorems are given in Section 4 along with
the proofs of the above lemmas.

Theorem 1. Under the stochastic ordering condition C_1 there is at most
a single triple $(v*,r*,s*)$ which simultaneously satisfies the system of
equations ⑤ with $r \leq s$.

Theorem 2. Equation (9a) has exactly one solution.

We note that regardless of whether or not (i) or (ii) obtains,
Theorem 2 insures that equation (9a) characterizes the optimal expected
return v^o in the class of policies in which testing is never permitted.

Theorem 3. _Under_ C_1 _and_ C_2 _and with_ $c_T > 0$, _condition_ (i) _of_
Lemma 3 _holds if and only if_ $T_1(v^o, v^o) \geq v^o$ _where_ v^o _is the unique_
solution to equation (9a).

Theorem 3 provides that testing is optimal for some value of
r if and only if at $r = v^o$ it is possible to do just as well as with
the best policy, say R^o, in the class which never permit testing, simply
by testing once and using R^o thereafter, for the best policy in this
class is in fact obtained from (9a), as we have seen.

To obtain the optimal policy in terms of the variables X and
Y, x^*, y^*, and x^o may be computed from r^*, s^*, and r^o, respectively,
using the transformation $x(r)$. Alternatively, the system of equations
6 and (9) and (10), can be formulated and solved in terms of the vari-
ables X and Y directly, since the one-to-one character of the
transformation $x(r)$ insures that the various uniqueness results given
above will carry over, as will the test for the form of the optimal
policy offered by Theorem 3.

4. Proofs

The function (3a) can be written in either of the forms

(11) $$T_1(r, v) = v + \int_v^\infty (1 - F_1(z|r)dz - c_T$$

or

(12) $$T_1(r, v) = r + \int_{-\infty}^v F_1(z|r)dz - c_T$$

These formulae may be verified by integrating by parts. Thus, for (11)
we find $v + z(1 - F_1(z|r))\big|_v^\infty + \int_v^\infty zf_1(z|r)dz - c_T$. To evaluate
$\lim_{z\to\infty} z(1 - F_1(z|r))$ we use $z(1 - F_1(z|r)) \leq \int_z^\infty zf_1(z|r)dx \to 0$ as $z \to \infty$

(since $\int_{-\infty}^{\infty} x f_1(x|r)dx$ exists and is finite.) Equation (12) is verified

in a similar manner using the fact that $r = \int_{-\infty}^{v} z f_1(z|r)dz + \int_{v}^{\infty} z f_1(z|r)dz$.

Lemma 1. To prove Lemma 1, we find from (11) that

$\partial T_1/\partial r = - \int_{v}^{\infty} \partial/\partial r F_1(z|r)dz > 0$, since under C_1, $\partial/\partial r F_1(z|r) < 0$, and

from (12) that $\partial T_1/\partial r = 1 + \int_{-\infty}^{v} \partial/\partial r F_1(z|r)dz < 1$, for the same reason.

Thus, $0 < \partial T_1/\partial r < 1$, as was to be shown.

Lemma 2. To prove Lemma 2, consider first equation (4). We need to show

that under C_2, and if $c_T > 0$, there is a solution to $T_1(r,v) = v$

for every fixed v. Inspection of (12) shows that for large positive

values of r, $T_1(r,v)$ will exceed v. For large negative values of r,

$T_1(r,v)$ will be below v. To see this, we choose the function β in

C_2 corresponding to δ_3 satisfying $|v|\delta_3 \leq c_T/2$ and $\delta_4 \leq c_T/2$, and

apply (1b) with r such that $\beta^0 = \beta(x(r)) \leq v$. Then

$$T_1(r,v) = v + v(F_1(v|r)-1) + \int_{v}^{\infty} z f_1(z|r)dz - c_T,$$

$$\leq v + |v|.|F_1(\beta^0|r)-1| + \int_{\beta^0}^{\infty} |z| f_1(z|r)dz - c_T,$$

$$\leq v + c_T/2 + c_T/2 - c_T = v,$$

From the continuity of T_1 there must be an intermediate value of r

for which $T_1(r,v) = v$.

Similarly, for equation (5), $T_1(s,v) \geq v-c_T \geq s$ for large

negative values of s. To show that $T_1(s,v) \leq s$ for large positive

values of s, the function α is chosen corresponding to $\delta_1, \delta_2 > 0$

for which $|v|\delta_1 \leq c_T/2$ and $\delta_2 \leq c_T/2$, and then s is selected so

that $\alpha^0 = \alpha(x(s)) \geq v$. Applying (1a),

$$T_1(s,v) = vF_1(v|s) + s - \int_{-\infty}^{v} zf_1(z|s)dz - c_T$$

$$\leq |v||F_1(\alpha^o|s)| + s + \int_{-\infty}^{\alpha^o} |z|f_1(z|s)dz - c_T$$

$$\leq c_T/2 + s + c_T/2 - c_T = s.$$

<u>Theorem</u> <u>1</u>. To prove Theorem 1, we note that by virtue of Lemma 1, it is only necessary to show that the system of equations (6), (7) and (8), subject to $r \leq s$, has a single solution in v, since then by this lemma r and s are uniquely determined. Let

$$\varphi(v) = v - [vF_1(r) + \int_{r}^{s} T_1(t,v)f_1(t)dt + \int_{s}^{\infty} tf_1(t)dt - c_S]$$

so that $\varphi(v) = 0$ is equivalent to (6). Differentiating φ subject to (7) and (8) we obtain

$$\varphi'(v) = 1 - [F_1(r) + vf_1(r)r' + T_1(s,v)f_1(s)s'$$
$$- T_1(r,v)f_1(r)r' + \int_{r}^{s} \partial/\partial v T_1(t,v)_1(t)dt$$
$$- sf_1(s)s'],$$

and on using (7) and (8),

$$\varphi'(v) = 1 - [F_1(r) + \int_{r}^{s} \partial/\partial v T_1(t,v)f_1(t)dt].$$

From (12) $\partial T_1/\partial v = F(v|r)$, so that

$$\varphi'(v) = 1 - [F_1(r) + \int_{r}^{s} F(v|t)f_1(t)dt].$$

But since $r \leq s$, $0 \leq \int_{r}^{s} F_1(v|t)f_1(t)dt \leq F_1(s) - F_1(r)$.
Thus,

$$0 < 1 - F_1(s) \leq \varphi'(v) \leq 1 - F_1(r) < 1$$

and $\varphi(v) = 0$ can only have one root.

__Theorem__ $\underline{2}$. We may prove Theorem 2 in a similar way, by differentiating

$\Psi(v) = v - [vF_1(v) + \int_v^\infty zf_1(z)dz - c_S]$. This gives $\Psi'(v) = 1 - F_1(v)$.

Hence $0 < \Psi'(v) < 1$, and $\Psi(v) = 0$ likewise only has one root. That

there is a root may be seen on integrating by parts as in (11) above.

Thus, $\Psi(v) = -\int_v^\infty (1-F_1(x))dx - c_S$ and $\Psi(v) \to -\infty$ as $v \to -\infty$, and

$\Psi(v) \to c_S > 0$ as $v \to +\infty$, so that $\Psi(v) = 0$ for some intermediate

value.

__Theorem__ $\underline{3}$. To prove Theorem 3, consider first the sufficiency part. We

have to show that if $T_1(v^o,v^o) \geq v^o$ where v^o satisfies

$v^o = v^o F_1(v^o) + \int_{v^o}^\infty zf_1(z)dz - c_S$, then there is at least one point r

such that if $R=r$ the optimal policy permits testing. Clearly $r = r^o = v^o$

is such a point since testing at $R=r^o$ and then using the best policy

without testing yields $T(v^o,v^o)$ and does at least as well as the latter

policy, which by Theorem 2 achieves exactly v^o.

Now we show that if there is a point r such that if $R=r$ testing

is optimal, then $T_1(v^o,v^o) \geq v^o$. Let $v^* \geq v^o$ be the optimal expected

return. Since max $\{v^*,r\}$ can be achieved by using the best of the two

choices, continuing optimally or stopping, the hypothesis is that for

some r, $T_1(r,v^*) \geq$ max $\{v^*,r\}$. Since $0 < \partial T_1/\partial r < 1$, this means that

$T_1(v^*,v^*) - v^* \geq T_1(r,v^*) - v^* \geq 0$ so that $T_1(v^*,v^*) \geq v^*$.

Now let $T_1(r,v;c) = vF_1(v|r) + \int_v^\infty zf_1(z|r) - c$; i.e., $T_1(r,v;c)$

is $T_1(r,v)$ with $c_T = c$ indicated explicitly. Obviously, $T_1(r,v;c)$

is strictly decreasing in c. Let (v_c,r_c,s_c) be the solution of ⑤

when $c_T = c$. We have then, by the above remark, that $T_1(v_c,v_c;c) \geq v_c$

for $v_c = v^*$. Let c increase from the given value c_T. Clearly,

$T_1(v_c, v_c; c)$ decreases with c, as does v_c, but from continuity considerations it is easily shown that at some point, say $c^o \geq c_T$, we will have

$$(14) \qquad T_1(v_{c^o}, v_{c^o}; c^o) = v_{c^o} .$$

This means $v_{c^o} = r_{c^o} = s_{c^o}$ and these are all equal to v^o, for on substituting $v_{c^o} = r_{c^o} = s_{c^o}$ in (6), (7) and (8) (read $T_1(r, v; c)$ instead of $T_1(r, v)$), (6) reduces to (9a), hence is satisfied by $v_{c^o} = v^o$, and (7) and (8) are satisfied, by application of (14). Moreover, this solution, $v_{c^o} = r_{c^o} = s_{c^o} = v^o$, is unique by Theorem 1. Thus we have $T_1(v^o, v^o; c^o) = v^o$. But $c_T \leq c^o$ hence $T_1(v^o, v^o; c_T) \geq v^o$ as was to be shown.

5. Selecting Many Possibilities

In this section, we will mention two possible applications of the model to related problems in which there is a cost constraint, and many possibilities are to be selected rather than just a single possibility as in Section 2. Concrete situations in which these problems might arise are to be found in the area of personnel selection and in connection with various biological selection problems, for example, drug screening.

Problem 1. From a very large population of possibilities, it is desired to select a fixed number N using a two stage testing procedure. The first test of a given possibility costs c_S^o. In order to apply the model, it is assumed the first test must be applied to each possibility. If the second test is used, it costs c_T^o. A measure of "true" worth W for the possibilities has a known (apriori) distribution

in the population of possibilities, and the regression of W on the
outcome of the first test, and on the outcome of the first and second
test combined, as well as the joint distribution of these regressions,
is known. The possibilities are taken at random from the population
and considered for selection one after another. The first test is applied
to each possibility, and after the outcome of the first test is known, the
possibility is either selected, rejected, or given the second test and
then either selected or rejected. The process is to be stopped as soon
as N possibilities are selected. The problem is to select the N
possibilities in such a manner as to make their expected total worth as
high as possible, subject to the constraint that the total testing cost
does not exceed a given testing budget C.

We cannot solve this problem exactly. However, we can determine
the policy which does well in items of the expected worth of the N pos-
sibilities selected, subject to the constraint that the expected total
testing cost is at most C. This policy has the advantage of being fixed;
that is, the same procedure is applied to each possibility, and if N is
large, say, forty or more, the actual cost of the policy will with high
probability not deviate from C by more than a small percentage amount
and the constraint will be approximately met. We need to assume that
$C > Nc_S^o$ in order for the approximate solution to make sense.

Consider the problem of selecting a single object as described
in Section 2, identifying X with the outcome of the first test, Y with
the outcome of the first and second test combined, and V with the
measure of worth W. However, instead of identifying the cost c_S with

c_S^O and c_T with c_T^O directly, we introduce a multiplier λ and with hypothetical costs $c_S = \lambda c_S^O$ and $c_T = \lambda c_T^O$, solve the optimization problem which consists of selecting a single possibility in such a manner as to maximize expected net return. This gives rise to an optimal policy depending on λ, say $r(\lambda)$. Associated with this policy will be its expected actual cost, say $Ec(\lambda)$ and the expected worth of the single possibility selected, say $Ev(\lambda)$. We now determine λ so that $Ec(\lambda) = C/N$ and select N possibilities using this policy repeatedly. It is assumed that such a λ exists and is positive. The expected total cost will then be C. The actual total cost in a specific instance will be the sum of N independent, identically distributed random variables, and we are assured by the strong law of large numbers that the ratio of the total cost to the expected cost approaches unity as N increases.

Now let Ev_i and Ec_i, $i=1,2,\ldots N$, be the expected return and expected cost, respectively, for selecting the ith possibility using any other procedure. We have

(15) $\qquad Ev(\lambda) - \lambda Ec(\lambda) \geq Ev_i - \lambda Ec_i$

by the optimality of $r(\lambda)$, and if $\sum_{i=1}^{N} Ec_i \leq C$, and $\lambda > 0$, we have $NEv(\lambda) \geq \sum_{i=1}^{N} Ev_i$ as required.

Problem 2. Suppose the situation is as in problem 1, with two stages of testing possible at costs $c_S^O > 0$ and $c_T^O > 0$, respectively, for the first and second tests. However, the requirement is to make the total expected worth of the possibilities selected as large as possible given a fixed budget, the number actually selected being unrestricted. Thus, the outcome will be random number N of random worths corresponding

to the possibilities selected. For large C, we can determine a "fixed" policy (see above) whose total cost is approximately C and does approximately as well as any other fixed policy.

As in problem 1, we first determine the optimal policy for a single search and evaluation problem with hypothetical costs $c_S = \lambda c_S^o$ and $c_T = \lambda c_T^o$, but now λ is chosen so that $Ev(\lambda) - \lambda Ec(\lambda) = 0$, where, as in problem 1, $Ev(\lambda)$ is the expected value of the single possibility selected using the optimal policy $r(\lambda)$ for that value of λ, and $Ec(\lambda)$ is the expected actual cost using this policy. Thus, in the hypothetical problem, λ is to be chosen so that $v*$ of Section 2 is zero. Since the model assumes positive costs, we suppose there is a positive λ which has this property.

The proposed policy consists of using $r(\lambda)$ over and over again until the testing budget C is exhausted, it being permitted, however, to use $r(\lambda)$ to complete selection of the last possibility even if the budget is exceeded while this is being done, but, of course, no new selections are started. Let r' be a given policy for selecting a single possibility and suppose the expected cost Ec' of the use of the policy is less than a given constant $c*$ and the variance of the cost, $\sigma^2 c'$, is less than a constant σ^2. Consider the class of policies formed by repeated use of any such policy r', with $Ec' < c*$ and $\sigma^2 c' < \sigma^2$, until the budget is exhausted in the above sense, that is, with selection of the possibility underway at the time the budget is exceeded being completed using r'. The proposed policy is approximately optimal in this class, in that for C sufficiently large, the proposed policy is almost certain to achieve a high proportion of the return of any element of the class. The proof of

this will be outlined, a number of technical details being omitted.[4]

For a given policy r' with $Ec' < c*$ for selecting a single possibility, let Ev' be the expected worth of the possibility selected under this policy. We have, then, by the optimality of $r(\lambda)$ in the hypothetical problem and by the choice of λ, that

$$(16) \qquad 0 = Ev(\lambda) - \lambda Ec(\lambda) \geq Ev' - \lambda Ec'.$$

If C is large relative to $Ec(\lambda)$, repeated use of $r(\lambda)$ until the budget is exhausted, implies, by virtue of the law of large numbers, the number N of selections actually made will be large. Thus, we will have $C \cong c_1^0 + c_2^0 + \ldots c_N^0 \approx NEc(\lambda)$, where $c_1^0 \ldots c_N^0$ are the random total costs incurred in making the N selections; more precisely, it can be shown that for some δ_1 whose absolute value is small relative to C with high probability, we will have $NEc(\lambda) = C + \delta_1$. Similarly, exhausting C by repeated use of any other policy r' means that for C sufficiently large, the corresponding number N' of selections will be large with high probability -- provided the policy has finite expected cost for selection a single possibility so that the law of large numbers may be applied. In fact, we will have $N'Ec' = C + \delta_2$ for some δ_2 which with high probability will have small absolute value relative to C for all r' such that $Ec' < c*$ and $\sigma^2 c' < \sigma^2$. Now, $N'(Ev' - \lambda Ec') \leq 0$ by (16), so $N'Ev' \leq \lambda N'Ec' = \lambda(C+\delta_2)$. Again, by (16), $NEv(\lambda) = \lambda NEc(\lambda) = \lambda(C+\delta_1)$,

---

[4] The interested reader can find the essential technical details worked out in Western Management Science Institute Working Paper No. 4, "Sequences of time variable games," available on request from the author.

so that $N'Ev' \leq NEv(\lambda) - \lambda\delta_1 + \lambda\delta_2$. The order of magnitude of the quantity $-\lambda\delta_1 + \lambda\delta_2$ is independent of N and the ratio of $NEv(\lambda)$ to $N'Ev'$ will in fact converge almost surely to a number $\alpha \leq 1$ as $C \to \infty$.

6. Computation of the Optimal Policy

The writer has been unable to find any interesting case where solutions to the system $\mathfrak{S}$ can be obtained in terms of elementary functions. Of course, the equations may be solved by numerical procedures in a specific instance. Some remarks on the computational problem in the case of the normal distribution[5] are perhaps in order:

If X, Y, and V in Section 2 have a joint normal distribution, it turns out that after making the obvious linear transformations, there are only three essential parameters in the model, of which two are the costs c_S and c_T. As was pointed out in Section 2, the distribution of X and the conditional distribution of the random variable $Z = E(V|X,Y)$, are all that count. These have a joint normal distribution which involves five parameters, the means and variances of X and Z and, say, their correlation, ρ. Choice of scale and origin for X and Z (or V), which are arbitrary in the model, eliminates the means and variances. For practical purposes, then, the computational problem reduces to tabling[6]

---

[5] Relevant tables as indicated below are being prepared for the case of the normal distribution.

[6] Using such tables, the problem of locating the multiplier λ referred to in problems 1 and 2 of Section 5 is easily solved by trial and error, entering the table with $c_S = \lambda c_S^0$ and $c_T = \lambda c_T^0$ until a value of λ is located such that the optimal expected cost (problem 1) or the optimal expected net return (problem 2) has the required property.

the optimal constants v^*, x^*, y^*, and v^o, and the expected costs under
the optimal policy, as a function of c_S, c_T, and ρ. Of course, a number
of other parameterizations are possible.

The joint normal distribution for X,Y, and V arises in the
situation where V has a known (apriori) normal distribution and the
decision maker is allowed to observe as the outcome of search a variable
ξ equal to V plus an error independent of V, while if he tests he is
allowed to observe a variable η equal to V plus another error inde-
pendent of V, the two errors leaving a known joint normal distribution.
Here we may identify ξ with X and η with Y.

This situation may be interpreted by saying the decision maker
learns about each V through a noisey channel, having the option to learn
more, at a cost, by using another noisey channel.

Another similar case giving rise to a joint normal distribution
is where V is the mean of a normal population, the population being
selected from a family of normal populations, all with the same variance,
in such a manner that V has itself a normal distribution. When a
population is selected for consideration, a sample of n_1 independent
observations on the population is first made. Further testing, if carried
out, results in another sample of n_2 independent observations. Here X
may be taken to be the mean of the first sample and Y to be the mean of the
second sample.

7. Acknowledgments

BIBLIOGRAPHY

1. Ashby, W. R., Design for a brain. New York, Wiley, 1952.

2. Bellman, Richard, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

3. Chow, Y. S., and Herbert Robbins, "A martingale system theorem and applications," in Proceedings of the fourth Berkeley symposium, Ed. J. Neyman, University of California Press, 1961.

4. Karlin, S., "Dynamic inventory policy with varying stochastic demands," Management Science, Vol. 6, No. 3, April, 1960, 231-258.

5. Lehmann, E. L., Testing statistical hypotheses, Wiley, New York, 1959.

6. MacQueen, J., and R. G. Miller, Jr., "Optimal persistence policies," Operations Research, Vol. 8, No. 3, May-June, 1960, 362-380.

7. Simon, Herbert A., "A behavioral model of rational choice," Chapter 14 in Models of Man, Wiley, New York, 1959.

8. Stigler, George J., "The economics of information," Journal of political economy, June, 1961, 213-225.

9. Toda, Masanao, "The design of a fungus-eater: A model of human behavior in an unsophisticated environment," Behavioral Science, Vol. 7, No. 2, April, 1962, 164-183.